

## ESTIMATION OF STOP LOSS PREMIUM IN FIRE INSURANCE

C. P. WELTEN  
Amsterdam, Holland

The estimation of stop loss premiums can be based on some knowledge about the distribution function of the sum of all claims in a year (assuming that the stop loss insurance relates to a period of one calendar year). Generally speaking there are two methods to obtain this knowledge about the distribution function.

- i. The first method is to construct a distribution function from data concerning:
  - a. the distribution function of the number of claims per year, taking into account the variability of the parameter(s) of this distribution function.
  - b. the distribution function of the insured sums.
  - c. the distribution function of partial claims.
  - d. the correlation between the insured sum and the probability of occurring of a claim.
  - e. the probability of contagion.
2. The second method is to derive a distribution function from the year's totals of claims over a long series of years, expressed in e.g. units of the totals of insured sums in that years.

In practice it is often difficult to find a useful basis to apply one of these methods. Data concerning the distribution function of the number of claims per year, of the insured sums, and of partial claims are mostly available, but often nothing is known about the correlation between the insured sum and the probability of occurring of a claim.

The second method is mostly not applicable because, if the year's totals of claims over a long series of, by preference recent, years are available, these data often turn out to be heterogeneous

or to be correlated with time. If, in that case, only the data of the most recent years are used, the number of these data is often a too small basis for the construction of a distribution function.

A special difficulty concerning the calculation of stop loss premiums is that we have to know as much as possible about events that happen only seldom. Often it is not easy to obtain that knowledge.

In the following we will describe a case of a stop loss cover in fire insurance. In this case, concerning a fire insurance company with a rather small number of insurances with relative high insured sums, the values

$$r_j = \frac{\text{total of claims, occurring in the year } j \times 10^4}{\text{total of insured sums in the year } j}$$

were available over a period of 45 years (1916-1960). As well in the numerator as in the denominator the reinsured parts of claims resp. reinsured parts of insured sums are not included. The data, necessary for application of the first method, were only partly available, and it was not possible to gather them as yet.

The coefficient of correlation between  $r_j$  and  $j$  proved to be  $-0,098$ . According to Fisher's  $z'$ -method the deviation of this value from zero is not significant, so that the hypothesis, that the values  $r_j$  have not a linear trend, can not be rejected. Further investigations have not given rise to the assumption that there is a non-linear trend of some importance, and indications for lack of homogeneity are not found. For this reasons we assumed that the 45 observed values  $r_j$  can be considered as a random sample from a homogeneous population.

An obvious method to find a useful distribution function from the values  $r_j$  is based on the assumption that this distribution function belongs to the wellknown class of Pearson-curves. To apply this method the central moments of the sample (the values  $r_1$  .....  $r_{45}$ ) are calculated first. These moments turn out to be:

$$\begin{array}{ll} m_1 = 2,256 & m_3 = 9,176 \\ m_2 = 4,205 & m_4 = 49,131 \end{array}$$

From these sample-moments unbiased estimates of the central moments of the population can be derived (see e.g. Cramèr, *Mathematical methods of Statistics*, p. 351). These estimates are in the present case:

$$\begin{aligned} S(\mu_1) &= 2,256 \\ S(\mu_2) &= 4,301 \\ S(\mu_3) &= 9,821 \\ S(\mu_4) &= 51,554 \end{aligned}$$

The Pearson-curve that has these central moments, proves to be a *B*-function:

$$f(r) = \frac{(r-p)^{m-1} \cdot (s-r)^{n-1}}{(s-p)^{m+n-1} \cdot B(m,n)} \quad (0,758 < r < 6,638)$$

where

$$\begin{aligned} p &= 0,758 \\ s &= 6,638 \\ m &= 0,134 \\ n &= 0,392 \end{aligned}$$

A first consideration showed already that 10 out of 45 values of the sample are situated below, and one above the interval of the *B*-function. Classification of the 45 values in the five intervals 0 - 1; 1 - 2; 2 - 3; 3 - 5 and > 5, and testing with the "expected" numbers in these intervals on the basis of the above mentioned *B*-function gives a value  $\chi^2 = 8,161$ , with four degrees of freedom; this does not lead to rejection of the considered frequency-function. Nevertheless it is remarkable that according to the *B*-function 11 out of 45 values have to fall in the interval > 5; in fact 7 values fall in this interval. It is therefore very doubtful, whether the above mentioned *B*-function gives a usefull representation of the "tail" of the actual frequency-function. Finally we mention about the *B*-function, that this function leads to a risk-premium for the stop-loss-cover (without safety-loading) ad 1,16 at a stop-loss-limit of 2,5, or a premium ad 0,79 at a limit of 3,5 (per 10.000 units of the total of insured sums).

The unacceptable results of application of the *B*-distribution have given rise to attempt another frequency-function namely the lognormal frequency-function.

$$f(r) dr = \frac{1}{\sqrt{2\pi\sigma}} \frac{1}{r} e^{-\frac{(\ln r - \mu)^2}{2\sigma^2}}$$

where

$$\mu = \frac{\sum_{j=1}^{45} \ln r_j}{45} = 0,394$$

$$\sigma^2 = \frac{\sum_{j=1}^{45} (\ln r_j - \mu)^2}{45} = 0,945$$

Classification of the observed values  $r_j$  in the same intervals as mentioned in the description of the  $B$ -distribution leads to  $\chi^2 = 4,148$  with four degrees of freedom; this figure does not lead to rejection of the lognormal frequency-function. The “expected” number of values in the interval  $> 5$  is 4,4; the observed number is 7. The deviation is indeed not significant, but nevertheless there is some suspicion against the lognormal frequency-function that it underestimates the “tail” of the actual frequency-function of  $r$ . Finally we mention about the lognormal frequency-function, that on the basis of this function the stop-loss-premium amounts to 0,72 at a stop-loss-limit of 2,5 and to 0,48 at a limit of 3,5 (per 10.000 units of the total of insured sums).

It must doubtlessly be possible to obtain a frequency-function, which fits the observed data of the sample reasonably well, also in the interval  $> 5$ . But then too it is not at all certain that such a frequency-function gives a good representation of the “tail” of the actual frequency-function; only if the sample should contain some hundreds of observed values it could be possible to make sure that the “tail” is well represented.

For that reason it is perhaps useful to look for an estimate of the stop-loss-premium, which is not based on any assumption about the analytical shape of the underlying distribution function. This can be done as follows.

If the random variable  $r$  (being  $10^4 \times$  the ratio between the total of claims in a year and the total of insured sums in that year)

has the frequency-function  $f(r)$ , a random variable  $h$  can be defined with the frequency-function  $F(h)$ :

$$\begin{aligned} F(h) &= 0 \quad \text{if } h < 0 \\ F(0) &= \int_0^0 f(r) dr \\ F(h) &= \int_0^h f(r) dr \quad \text{if } h > 0 \end{aligned}$$

Then the mean  $P$  of  $h$  is the risk premium for the stop loss cover at the limit  $G$ ;  $\sigma P$ , being the standard deviation of  $h$ , is the standard-deviation of the yearly stop-loss-risk. The 45 observed values  $h_j$  can be derived from the values  $r_j$  according to:

$$\begin{aligned} h_j &= 0 \quad \text{if } r_j \leq G \\ h_j &= r_j - G \quad \text{if } r_j > G \end{aligned}$$

The values  $h$  can be considered as observed values of the random variable  $h \sim F(h)$ . From this follows:

1. that  $\bar{P} = \frac{\sum_{j=1}^{45} h_j}{45}$  is an unbiased estimate of the stop loss premium  $P$ .
2. that  $S_{\bar{P}}^2 = \frac{\sum_{j=1}^{45} (h_j - \bar{P})^2}{44}$  is an unbiased estimate of  $\frac{44}{45} \sigma_P^2$
3. that  $\sigma_{\bar{P}}^2$ , being the expected value of  $(\bar{P} - P)^2$ , is  $\frac{\sigma_P^2}{45}$ .

Therefore the estimate of the standard deviation of the estimate  $\bar{P}$  can be calculated as  $\sigma_{\bar{P}}^2 = \frac{1}{44} \cdot S_{\bar{P}}^2$

In the present case the estimates  $\bar{P}$  and  $\sigma_{\bar{P}}^2$  turn out to be:

$$\text{at a limit } G = 2,5; \quad \bar{P} = 0,77; \quad \sigma_{\bar{P}} = 0,22$$

$$\text{at a limit } G = 3,5; \quad \bar{P} = 0,51; \quad \sigma_{\bar{P}} = 0,15$$

In practice the stop loss premium can be calculated as

$$\bar{P} + \alpha \sigma_{\bar{P}} + \beta \sigma_{\bar{P}}$$

In this formula the value of  $P + \alpha\sigma_P$  can be considered as the upper limit of a confidence interval, which contains the stop loss premium  $P$  with a probability, for which Gauss'inequality gives a low estimate. The value  $\alpha\sigma_P$  is a security-loading.

## CONCLUSION

The abovementioned case demonstrates that in calculating stop-loss-premiums a cautious choice of the underlying distribution function is necessary, and that it is perhaps better to make no assumption concerning the shape of that function. It proves—at least in some cases—to be possible to obtain a “quantified uncertainty” about the stop loss premium without such assumption.

## APPENDIX

$j$	$v_j$	$j$	$v_j$
1	0,66	24	0,58
2	6,47	25	0,26
3	0,45	26	1,36
4	4,51	27	1,98
5	6,48	28	4,90
6	2,03	29	3,36
7	4,69	30	1,90
8	0,26	31	0,97
9	0,67	32	1,40
10	3,88	33	6,09
11	2,50	34	0,81
12	0,30	35	0,77
13	0,95	36	0,49
14	1,19	37	2,07
15	2,78	38	0,76
16	0,51	39	5,37
17	1,28	40	0,79
18	7,26	41	1,70
19	0,44	42	0,93
20	1,01	43	5,11
21	1,59	44	1,11
22	0,88	45	1,61
23	6,42		